Paper ID #1665

Multi-Level Region Matching for Fine-Grained Sketch-Based Image Retrieval

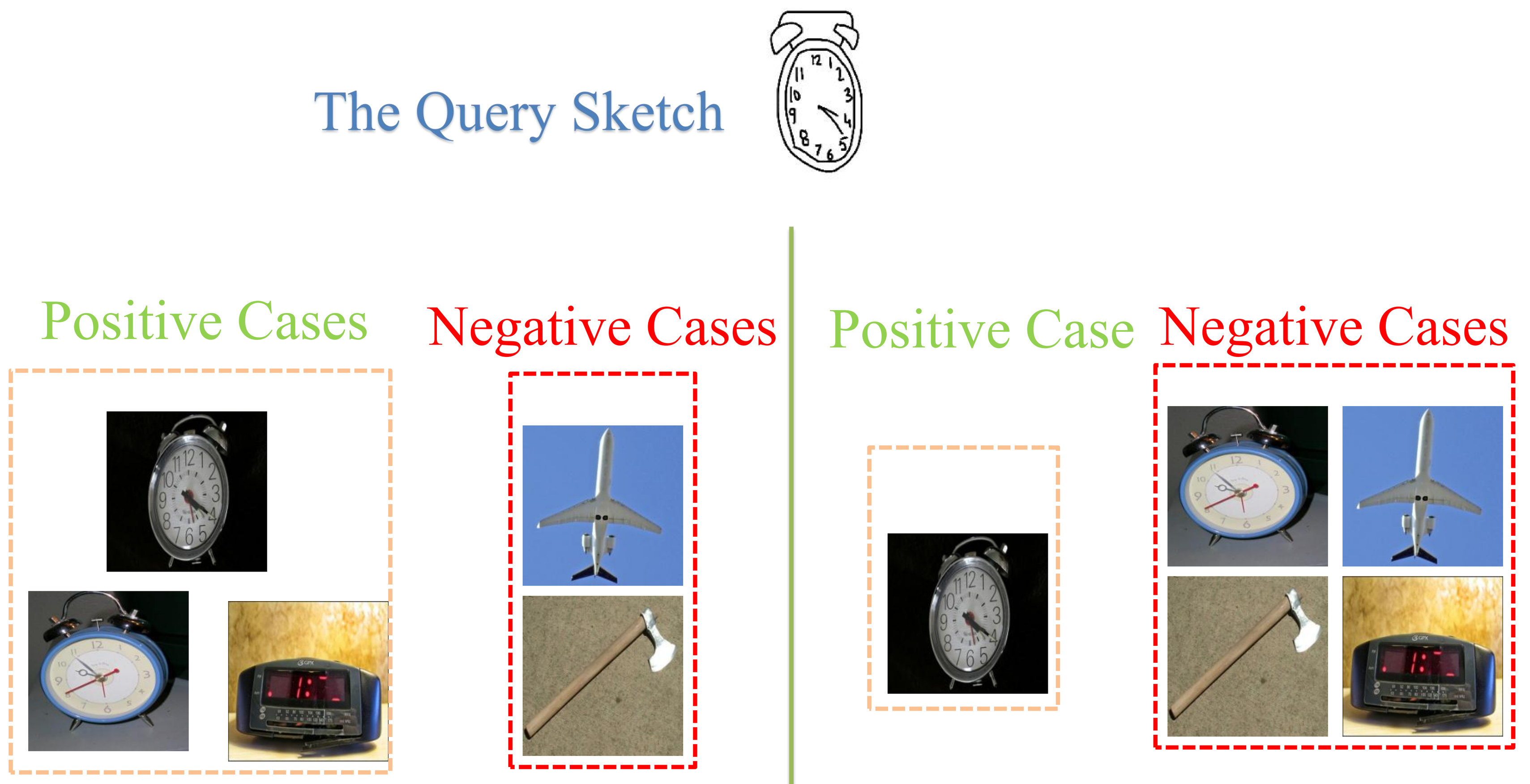
Zhixin Ling, 1069066484@qq.com, Fudan University

Zhen Xing, zxing20@fudan.edu.cn, Fudan University

Jiangtong Li, keep_moving-Lee@sjtu.edu.cn, Shanghai Jiao Tong University

Li Niu*, ustcnewly@sjtu.edu.cn, Shanghai Jiao Tong University

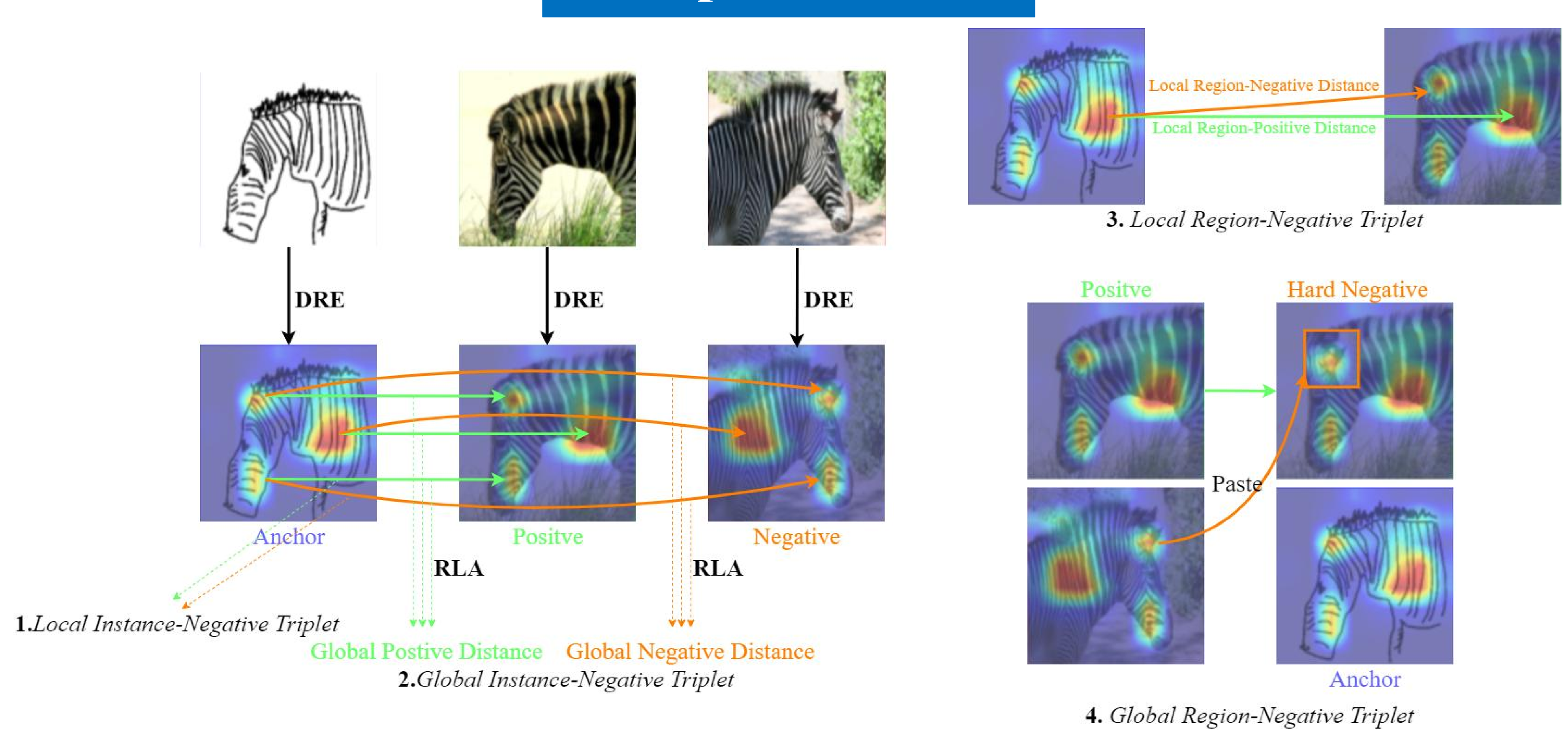
Task Setting



Coarse-Grained Sketch-Based Image Retrieval (**CG-SBIR**)

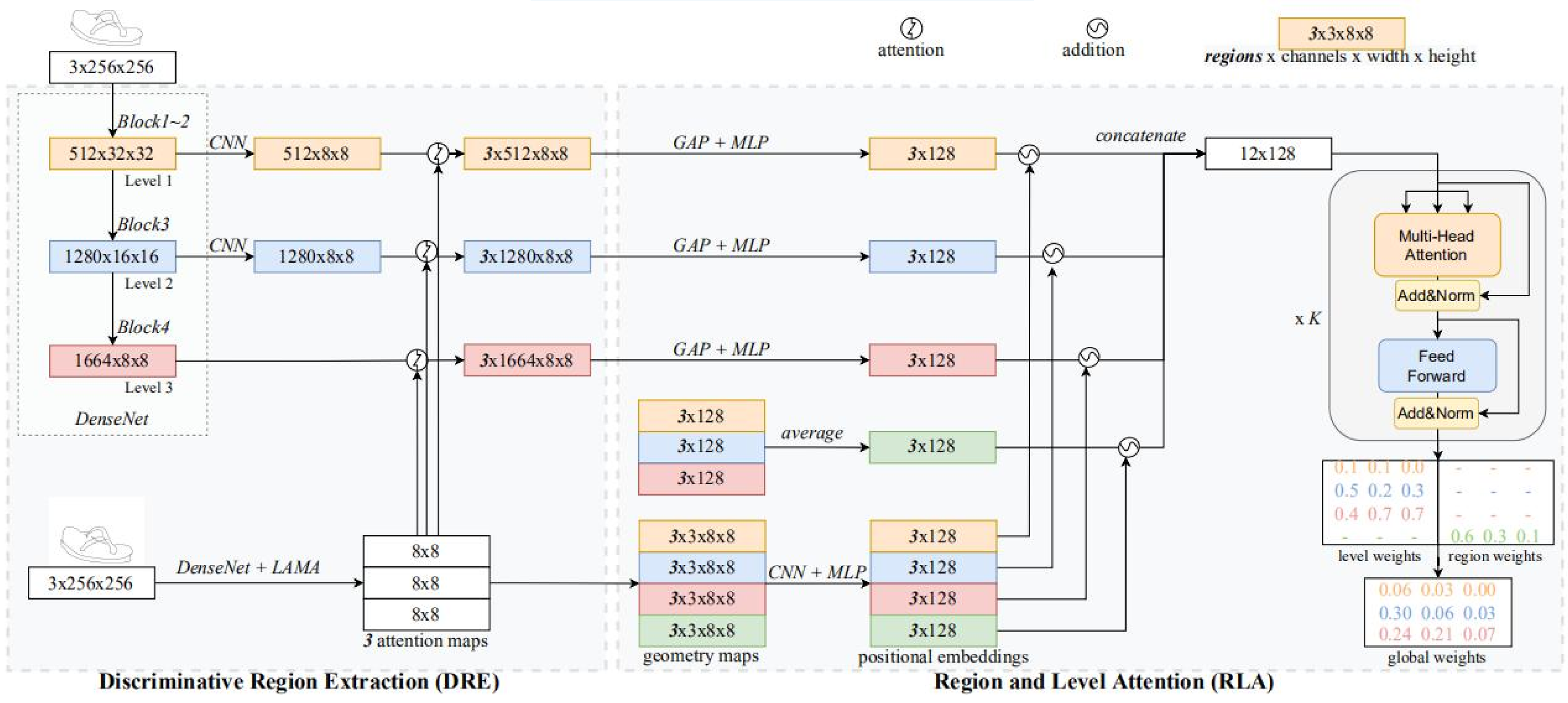
Fine-Grained Sketch-Based Image Retrieval (**FG-SBIR**)

Triplet Losses



Triplet loss 1: $\mathcal{L}_{gtrp-in}$ targets at semantic correspondence between paired regions.
Triplet loss 2: $\mathcal{L}_{gtrp-rn}$ targets at global matching distance.
Triplet loss 3: $\mathcal{L}_{ltrp-in}$ targets at semantic distinctiveness across unpaired regions.
Triplet loss 4: $\mathcal{L}_{ltrp-rn}$ targets at hard negative samples.

Framework



Overview of the proposed MLRM

- In DRE, we propose a LAMA structure to extract different attention maps to attend multi-level CNN feature maps.
- In RLA, we adopt a transformer-based attentive matching module to obtain attention weights for different regions and levels.
- At last, we aggregate region/level-wise distances by weights as a retrieval distance.

LAMA *v.s.* CAMA

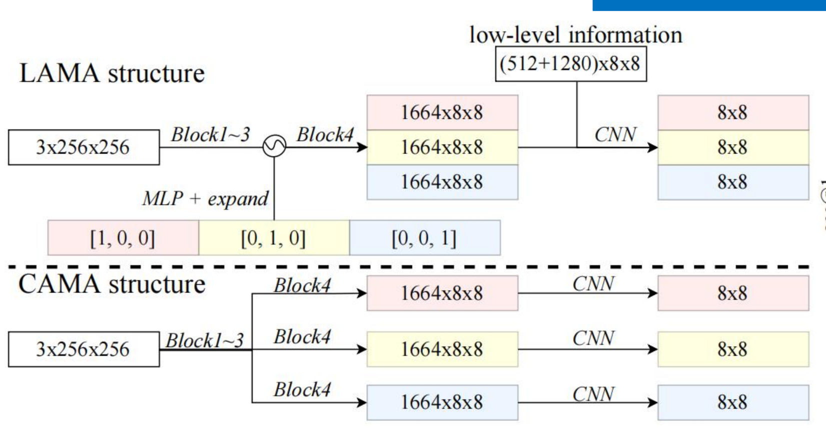


Figure 2: LAMA and CAMA structure.

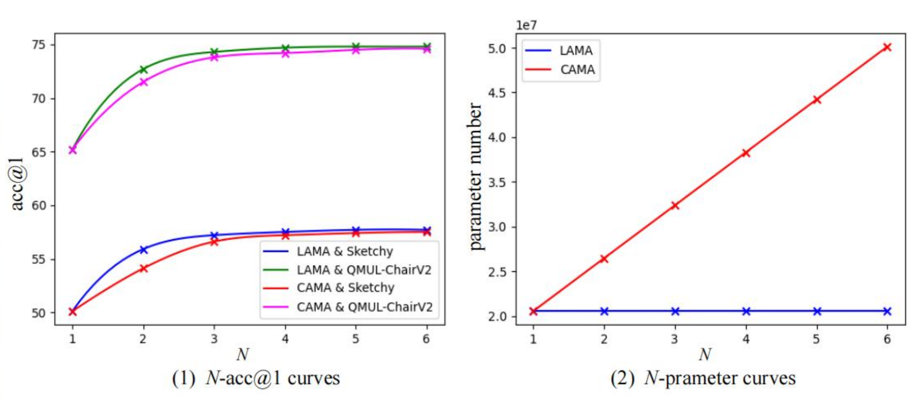
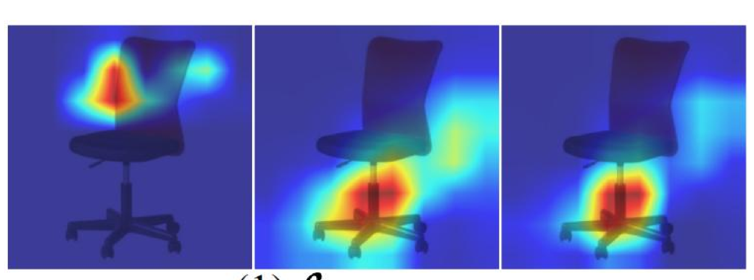


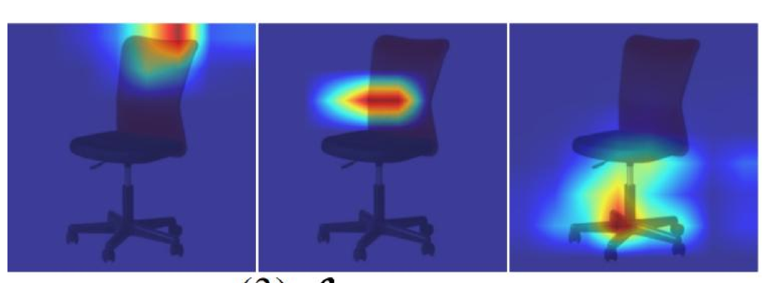
Figure 4: LAMA and CAMA quantitative comparison.

$$\mathcal{L}_{ovl-cama} = \frac{1}{N}\sum_{x,y} M_1 \odot M_2 \odot \cdots \odot M_N,$$

$$\mathcal{L}_{ovl-lama} = \frac{1}{N\times N}\sum_{r\le N}\sum_{x,y} M_r \odot MaxPool(\prod_{r'\ne r, r'\le N} M_{r'})$$



(1) $\mathcal{L}_{ovl-cama}$



(3) $\mathcal{L}_{ovl-lama}$

- Inspired by CAMA, we propose LAMA to extract discriminative regions.
- LAMA merges different network branch copies into one, saving a large number of parameters when improving model performance.
- LAMA adopts an improved overlapping penalty, learning better geometrically discriminative regions.

Experiments

	Sketchy (%)	QMUL-ChairV2 (%)	QMUL-ShoeV2 (%)	QMUL-Chair (%)	QMUL-Shoe (%)
Song *et al.* [32] (CVPR '16)	-	-	-	78.4	50.4
GN Triplet [29] (TOG '16)	37.1	-	-	-	-
SaN Triplet [42] (CVPR '16)	36.2	56.6	30.9	72.2	52.2
Quadruplet [30] (ACM MM '17)	42.2	-	-	-	-
DSSA [33] (ICCV '17)	-	-	33.7	81.4	61.7
Radenovic *et al.* [27] (ECCV '18)	-	-	-	85.6	54.8
DCCRM [40] (PR '19)	46.2	-	-	-	-
TC-Net [19] (ACM MM '19)	40.8	65.3	40.2	95.9	63.5
Bhunia *et al.* [5] (CVPR '20)	-	(89.7)	(79.6)	-	-
Pang *et al.* [26] (CVPR '20)	-	-	36.5	96.0	56.5
Bhunia *et al.* [3] (CVPR '21)	-	60.2	39.1	-	-
LA [37] (ACM MM '21)	43.1	64.8	42.3	93.8	57.4
DLA [37] (ACM MM '21)	54.9	69.2	50.2	**99.0**	**79.1**
Zhang *et al.* [43] (PR '22)	-	-	-	84.4	65.7
AE-Net [7] (PR '22)	46.0	-	-	-	-
Bhunia *et al.* [4](CVPR '22)	-	64.8	43.7	-	-
MLRM (ours)	**57.2**	**74.3(98.2)**	**50.4(87.9)**	**99.0**	67.0

Table 1: acc@1(acc@10) comparison with previous works.

		TC-Net[19]	LA [37]	DLA [37]	MLRM (ours)
QMUL-ChairV2	Time (s)	5.3	27.5	236.7	11.8
	acc@1 (%)	65.3	64.8	69.2	74.3
Sketchy	Time (s)	8.2	46.8	639.3	14.1
	acc@1 (%)	40.8	43.1	54.9	57.2

Table 2: Retrieval time comparison using the same GPU.

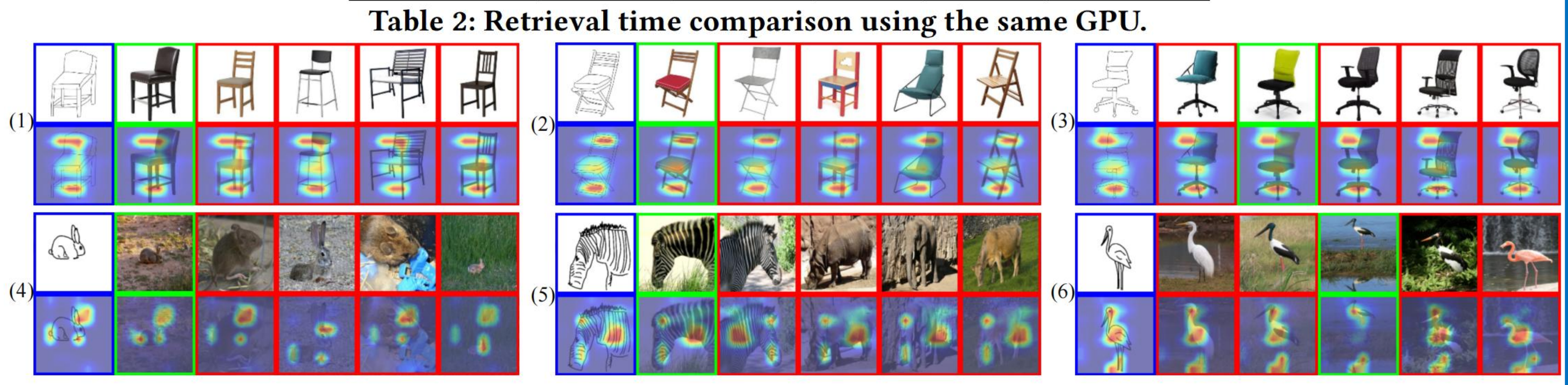


Figure 6: Top-5 retrieval visualization on QMUL-ChairV2(row (1)-(3)) and Sketchy(row (4)-(6)). The sketches bordered in blue are queries. The images bordered in green/red are positive/negative cases.

- Our MLRM achieved SOTA on all datasets except QMUL-Shoe, on which MLRM is the second best.
- Our MLRM does not introduce much extra computation overhead.
- Our MLRM can well extract both geometrically and semantically discriminative regions.

Conclusion

- To establish fine-grained correspondence between sketches and im_x0002_ages, we propose Multi-Level Region Matching (MLRM).
- MLRM consists of DRE that generates discriminative regions and RLA to obtain attention weights for different regions and levels.
- Comprehensive experiments have demonstrated that MLRM achieves SOTA acc@1 at the cost of a relatively low computation overhead.

Acknowledgements

The work is supported by Shanghai Municipal Science and Technology Major Project, China (2021SHZDZX0102), and Shanghai Municipal Science and Technology Key Project (Grant No. 20511100300), and National Science Foundation of China (61902247).